Final Report AFOSR FA9550-06-1-0408

Studies of single biomolecules, DNA conformational dynamics, and protein binding

Andreas Hanke

Department of Physics and Astronomy,

University of Texas, 80 Fort Brown, Brownsville

Abstract

While the Watson-Crick double-strand is the thermodynamically stable state of DNA in a wide range of temperature and salt conditions even at physiological conditions local denaturation bubbles may open up spontaneously due to thermal activation. By rising the ambient temperature, titration, or by external forces in single molecule setups bubbles proliferate until full denaturation of the DNA. Based on the Poland-Scheraga model we investigate both the equilibrium transition of DNA denaturation and the dynamics of the denaturation bubbles with respect to recent single DNA chain experiments for situations below, at, and above the denaturation transition. We also propose a new single molecule setup based on DNA constructs with two bubble zones to measure the bubble coalescence and extract the physical parameters relevant to DNA breathing. Finally we consider the interplay between denaturation bubbles and selectively single stranded DNA binding proteins.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
11-07-2008	FINAL	APRIL 2005 - APRIL 2008

4. TITLE AND SUBTITLE

(CONGRESSIONAL - ESP CODE: VU) STUDIES OF SINGLE BIOMOLECULES: DNA CONFORMATIONAL DYNAMICS AND PROTEIN BINDING

5a. CONTRACT NUMBER

5b. GRANT NUMBER

FA9550-06-1-0408

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

HANKE, ANDREAS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UNIVERSITY OF TEXAS AT BROWNSVILLE
DEPARTMENT OF PHYSICS
80 FORT BROWN, TX 78520

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AF OFFICE OF SCIENTIFIC RESEARCH
875 NORTH RANDOLPH STREET ROOM 3112
ARLINGTON VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

While the Watson-Crick double-strand is the thermodynamically stable state of DNA in a wide range of temperature and salt conditions even at physiological conditions local denaturation bubbles may open up spontaneously due to thermal activation. By rising the ambient temperature, titration, or by external forces in single molecule setups bubbles proliferate until full denaturation of the DNA. Based on the Poland-Scheraga model we investigate both the equilibrium transition of DNA denaturation and the dynamics of the denaturation bubbles with respect to recent single DNA chain experiments for situations below, at, and above the denaturation transition. We also propose a new single molecule setup based on DNA constructs with two bubble zones to measure the bubble coalescence and extract the physical parameters relevant to DNA breathing. Finally we consider the interplay between denaturation bubbles and selectively single stranded DNA binding proteins.

15. SUBJECT TERMS

Single-molecule biophysics, DNA structure and conformational dynamics, DNA protein binding

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			ANDREAS HANKE
U	U	U	SAR	32	19b. TELEPHONE NUMBER *(Include area code)* 956-882-6682

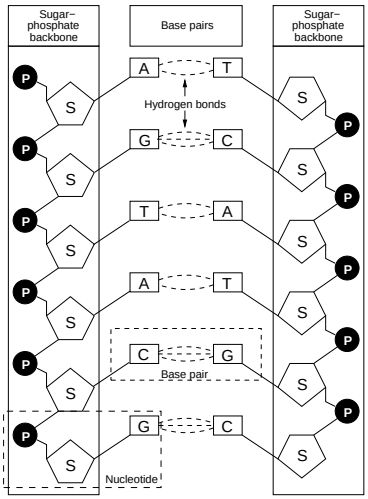

FIG. 1: Ladder structure of DNA showing the Watson-Crick bonding of the bases A, T, G, and C which are suspended by a sugar-phosphate backbone. Each phosphate carries a negative charge. The longitudinal distance between adjacent base pairs is 3.43 Å while approximately 10.5 base pairs are needed to form a complete helical turn. Under normal salt conditions the persistence length of double stranded DNA is approximately 50 nm, the hard core diameter is approximately 2 nm. Locally (i.e., for lengths shorter than the persistence length), DNA appears thin and stiff while on longer scales it can be perceived as a flexible polymer. The length of a single DNA molecule varies from several μm of viral DNA over several mm in bacteria up to many cm in eukaryotic cells.

I. INTRODUCTION

Deoxyribonucleic acid (DNA) is the molecule of life, encoding the complete genetic information of an entire organism. This information is kept in terms of the four letter alphabet comprised by Adenine, Guanine, Cytosine, and Thymine. The genetic code is stabilised by base pairing through hydrogen bonding that creates two complementary strands subject to the key-lock principle. This way it is made sure that exclusively AT and GC nucleotides pair. Within this ladder structure (Fig. 1) the bases and thus the genetic code are protected against unwanted action of chemicals and proteins. The three-dimensional structure of DNA is the famed Watson-Crick double-helix, the equilibrium structure of DNA within a broad range of salt and temperature conditions. Sufficiently close to physiological conditions the typical conformation of double-stranded DNA is the B form with a pitch of 3.4 Å between successive base pairs and approximately 10.5 base pairs needed to form one complete turn of

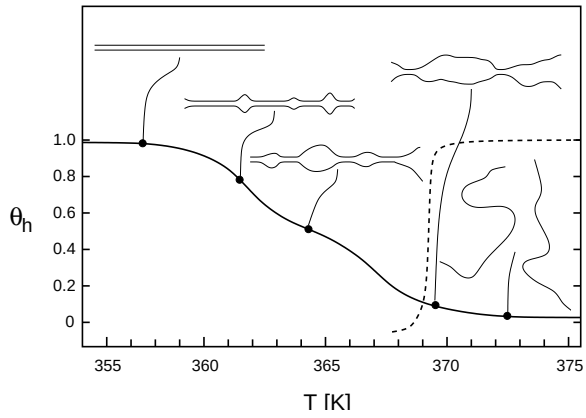

FIG. 2: Thermal denaturation of double stranded DNA: Fraction θ_h of double-helical domains within the DNA as a function of temperature. Schematic representation of $\theta_h(T)$, showing the increased formation of bubbles and unzipping from the ends, until full denaturation has been reached. Note that bacterial DNA is predominantly circular so that no end effects occur. Also viral DNA circularises once injected into a host cell.

the helix. This thermodynamic stability apart from hydrogen-bonding between paired bases is mainly effected by base-stacking between nearest neighbour pairs of base pairs [1–6].

By temperature increase or variation of the pH (titration with acid or alkali) double-stranded DNA progressively denatures. The comparatively stiff DNA double-strand (persistence length *circa* 50 nm) is thereby interrupted by emerging zones of flexible single-strand (persistence length circa 1 to a few nm). These so-called DNA bubbles then grow and merge until the double-strand is fully molten (Fig. 2). This is the helix-coil transition. The melting temperature T_m is experimentally defined as the temperature at which half of the DNA molecule has undergone denaturation [3, 7, 8]. Typically, the denaturation starts in regions rich in the weaker AT base-pairs, and subsequently moves to zones of increasing GC content. The occurrence of zones of different stability within the genome was shown to be relevant when separating coding from non-coding regions [9, 10] and is believed to be related to DNA function, for instance the occurrence of weak regions (*e.g.* the TATA motif) at transcription initiation points.

Albeit rare, already at room temperature thermal fluctuations cause opening events of small intermittent denaturation bubbles [11]. The size of these bubbles fluctuates by stepwise zipping and unzipping of the base pairs at the zipper forks where the bubble con-

nects to the intact DNA double-strand (bubble breathing). Initiation of a bubble in a stretch of double-strand requires the crossing of a free energy barrier F_s of approximately 8 kcal/mol (some 10 k_BT at physiological temperature) corresponding to a Boltzmann factor $\sigma_0 = \exp(-F_s/k_BT) \sim 10^{-5\ldots-3}$. σ_0 is often referred to as the cooperativity factor. Once formed below the melting temperature T_m a bubble will eventually zip close. Above T_m, a bubble will preferentially stay open and, if unconstrained, grow in size until it merges with other denaturation bubbles, eventually leading to full denaturation. Constraints against such full unzipping could, for instance, be the build-up of twist in smaller DNA-rings [12], the highly positively supercoiled state (linking excess) in the DNA of extremophile bacteria existing at high temperatures in deepsea vents [13], or the chemical connection of the two strands by short bulge-loops, compare Ref. [14]. In heteropolymer DNA mechanical stretching experiments show that even at the end of the overstretching transition and beyond the two strands do not separate completely [15–17] but are still held together by isolated GC-rich regions along the chain with average distance of a few hundreds of base pairs [16]. These GC-rich regions break only at a much larger force than the melting force F_m of the overstretching plateau.

Biologically the physical conformations of DNA molecules are recognised to be of inalienable relevance for its function, see, for instance, the review [18] and references therein. In particular, the existence of intermittent though infrequent bubble domains is important as the opening up of the Watson-Crick base pairs by breaking of the hydrogen bonds between complementary bases disrupts the helical stack. The associated flipping out of the ordered stack of the unpaired bases allows the binding of specific chemicals or proteins, that otherwise would not be able to access the reactive sites of the bases [3, 6, 7, 11]. Indeed there exists a competition of time scales between the survival of DNA-bubbles and the binding kinetics of selectively single-stranded DNA binding proteins [19–21]. An important aspect to the biological function of DNA it is believed that DNA-breathing assists transcription initiation [22–25], see below. Altogether it appears fair to say that the quantitative knowledge of the energetics of the denaturation as well as the dynamics of bubbles is imperative to a better understanding of genomic biochemical processes. Additionally DNA denaturation is a fine example of a well defined and chemically stable system whose physical properties can be probed in detail on the level of single molecules. DNA is therefore studied from both viewpoints biological physics with respect to DNA's rôle in biochemical processes and

statistical physics for which DNA provides an ideal system to study quantitatively polymer models.

In what follows we will base our analysis on the Poland-Scheraga free energy model treating the DNA molecule as an Ising-type system of a sequence of 'spins' with open or closed states plus a non-local term that takes care of polymeric effects within denaturation bubbles made up of highly flexible DNA single-strand. A prominent alternative description of DNA denaturation and breathing is the Peyrard Bishop Dauxois (PBD) model [26, 27] based on the set of Langevin equations [28]

$$m\frac{d^2y_n}{dt^2} = -\frac{dV(y_n)}{dy_n} - \frac{dW(y_{n+1},y_n)}{dy_n} - \frac{dW(y_n,y_{n-1})}{dy_n} - m\gamma\frac{dy_n}{dt} + \xi_n(t). \quad (1)$$

Here, $V(y_n) = D_n\left[\exp(-a_n y_n) - 1\right]^2$ is a Morse potential for the hydrogen bonding, D_n and a_n assuming two different values for AT and GC bps; $W(y,y') = \frac{k}{2}\left[1+\rho\exp\{-\beta(y+y')\}\right](y-y')^2$ is a nonlinear potential to include bp-bp stacking interactions between adjacent bps y and y'. The parameters k, ρ, β, γ, and m are invariant of the sequence. The equation is driven by the thermal noise $\xi_n(t)$. Usually, the stochastic equations (1) are integrated numerically [28]. Due to its formulation in terms of a set of Langevin equations, the DPB model is very appealing, and it is a useful model to study some generic features of DNA denaturation. Its disadvantage is that somewhat arbitrary values for the model parameters need to be chosen while (apart from the characteristic time scale) all parameters in the Poland-Scheraga model are available from a large body of experiments.

We first address the denaturation transition at equilibrium both in absence and presence of an external stretching force. Subsequently we will present two model approaches to the breathing dynamics of a single denaturation bubble. In Section 4 we discuss the coalescence dynamics of two DNA bubbles. Finally, in Section 5 we address the coupling of the breathing dynamics of a DNA bubble with the binding/unbinding of proteins that specifically bind to single-stranded DNA.

II. DNA DENATURATION IN PRESENCE OF A MODEST STRETCHING FORCE

A convenient method to treat the denaturation transition is to consider the chain in the grand canonical ensemble in which the total number N of bps and the end-to-end vector $\mathbf{L}$ fluctuate. The partition function in $d = 3$ of the DNA chain under external forcing with force F in x direction becomes [29]

$$\mathcal{Z}(z, F) = \sum_{N=1}^{\infty} \int d^3 L \mathcal{Z}_{\rm can}(N, \mathbf{L}) z^N \exp(\beta F L_x) \tag{2}$$

with $\beta = 1/(k_B T)$. $\mathcal{Z}_{\rm can}(N, \mathbf{L})$ is the canonical partition function of a chain of N bps with fixed end-to-end vector $\mathbf{L}$, z is the fugacity, and L_x the x-component of $\mathbf{L}$ (Fig. 3). Assuming that bound segments and bubbles are independent, $\mathcal{Z}$ factories:

$$\mathcal{Z}(z, F) = \Omega_e + \Omega_e \left\{ \sum_{n=0}^{\infty} [B\Omega]^n \right\} B\Omega_e = \Omega_e + \frac{\Omega_e^2 B}{1 - \Omega_e B}. \tag{3}$$

The alternating sequence of bound segments and bubbles with weights B and Ω in Eq. (3) is complemented by the weight Ω_e of an open end unit at both ends of the chain. We assume that only one strand of the end unit is bound to the, say, magnetic bead, while the other strand is moving freely. Thus the first term on the right hand side of Eq. (3) denotes the two unbound single strands of completely denatured DNA; here we assume that one of the two strands is still attached between origin $\mathcal{O}$ and end point $\mathbf{L}$, being subject to the stretching force F.

A bound segment with $k = 1, 2, \ldots$ bps is modelled as a rigid rod of length ak where $a = 0.34\,\mathrm{nm}$ is the length of a bound bp in B-DNA [30]. Here we assume a homopolymer with binding energy $E_0 < 0$ per base pair. However, we assume perfect matching throughout the transition such that in a denaturation bubble both single stranded arches carry equal length. This assumption is in line with above remark that due to stable GC-rich islands in the structure during a force-induced denaturation the sequence of separated denaturation bubbles and intact double-strand persists to much larger forces than the melting force F_m of the overstretching plateau [16]. The statistical weight of a segment with fixed number k and fixed orientation is then ω^k with $\omega = \exp(\beta\varepsilon)$ and $\varepsilon = -E_0 > 0$. Assuming that k fluctuates with fixed fugacity z, and rotates around one end while subject to the force F

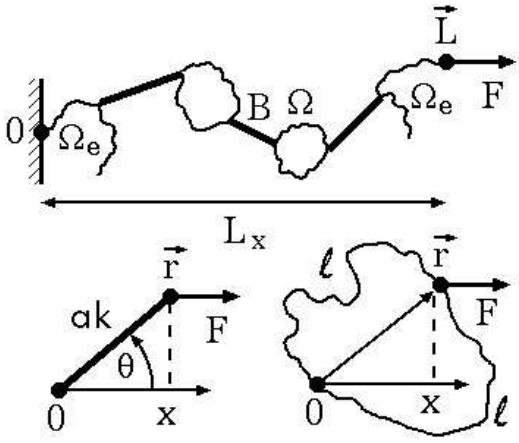

FIG. 3: Stretched DNA in the PS model with bound segments B and denatured loops Ω. The DNA is attached between $\mathcal{O}$ and $\mathbf{L}$ and subject to the stretching force F in x-direction. Perfect matching in heterogeneous DNA requires both arches of a loop to have equal length ℓ.

(Fig. 3), the statistical weight of the segment for fixed z and F becomes [29]

$$B(z,\omega,F)=\frac{1}{2y}\ln\left(\frac{1-\omega z e^{-y}}{1-\omega z e^{y}}\right),\ y\equiv\beta F a. \tag{4}$$

At $F=0$, $B(z,\omega,0)=\omega z/(1-\omega z)$ as found previously for free DNA [31].

A denatured loop is considered as a closed random walk with 2ℓ monomers, corresponding to ℓ broken bps. The loop starts at $\mathcal{O}$ and visits the point $\mathbf{r}$ after ℓ monomers (Fig. 3). The number of configurations of such a loop becomes

$$\Omega(\ell,\mathbf{r})=C_0(2\ell)p_\ell(\mathbf{r}) \tag{5}$$

where $C_0(2\ell)$ counts the configurations of a loop of length 2ℓ starting at $\mathcal{O}$ and $p_\ell(\mathbf{r})$ is the probability that the loop visits $\mathbf{r}$ after ℓ monomers. For an ideal random walk in $d=3$, $C_0(2\ell)\sim\mu^{2\ell}\ell^{-3/2}$ (μ is the connectivity constant[89]) and $p_\ell(\mathbf{r})\sim\mathcal{R}^{-3}\exp[-\lambda(r/\mathcal{R})^2]$ where $\lambda>0$, $r=|\mathbf{r}|$, and $\mathcal{R}=b\ell^{1/2}$ is the scaling length of the walk. The coefficient b is proportional to the persistence length. Thus, $\Omega(\ell,\mathbf{r})\sim s^\ell\ell^{-3}\exp[-\lambda(r/\mathcal{R})^2]$ where $s=\mu^2$. We assume that $\mathbf{r}$ moves freely and is subject to the force F in the positive x-direction. The weight of an ideal random loop for fixed ℓ and F is given by the Gaussian integral

$$\Omega(\ell,F)=\int d^3r\,\Omega(\ell,\mathbf{r})\,e^{\beta F x}=As^\ell\ell^{-c}\exp(\alpha y^2\ell) \tag{6}$$

where A is an amplitude proportional to the cooperativity factor σ_0, $c=3/2$, and $\alpha=b^2/(4\lambda a^2)$.

Generally, the loop free energy in both presence and absence of the external forcing is of the power-law form

$$\Omega \simeq \ell^{-c}. \tag{7}$$

For free DNA it was found that the nature of the denaturation transition is determined by the value of the critical exponent c: for $c \leq 1$ there is no phase transition in the thermodynamic sense; for $1 < c \leq 2$ the transition is second order, and for $c > 2$ it is first order [31–33]. One finds $c = 3/2 < 2$ if the loops are ideal random walks. Self-avoiding interactions within a loop modify this value to $c = 3\nu = 1.76$ with $\nu = 0.588$ in $d = 3$. In both cases the transition is second order. Including self-avoiding interactions between denatured loops and the rest of the chain was found to produce $c = 2.12 > 2$, driving the transition to first order [9, 31, 34]. These results suggest that the inclusion of self-avoiding interactions generally shifts the loop exponent c to larger values, possibly effecting a change of the transition from second to first order.

Using scaling arguments in the presence of self-avoiding interactions within a loop we find a modified expression for the statistical weight [29]

$$\Omega(\ell, F) = A s^{\ell} \ell^{-c} y^{1/(2\nu)-1} \exp\left(\alpha y^{1/\nu} \ell\right) \tag{8}$$

for $\kappa = \beta b F \ell^{\nu} \to \infty$ and with the new loop exponent in $d = 3$,

$$c = 4\nu - 1/2 = 1.85. \tag{9}$$

Thus, in the presence of self-avoiding interactions within a denatured loop and $F > 0$ the transition remains second order, but moves closer to first order compared to free DNA (with $c = 3\nu = 1.76$ obtained within the same approach). In the Gaussian limit the same result obtains as in the absence of the force corresponding to the ideal Hookean chain behaviour of a phantom chain.

Within this formalism it is also possible to obtain the force-extension behaviour of the chain as well as the temperature-force phase diagram, see Fig. 4. The shape of the transition line $f_m(t)$ depends on A, α, and s. Fig. 4a shows $f_m(t)$ for $A = 1$, $\alpha = 1$, and $s = 5$ for the case that denatured loops are ideal random walks ($\theta = 0$, $\nu = 1/2$). The transition line for a more realistic value $A \ll 1$ is also shown (here $A = 0.01$). The line $f_m(t)$ separates a finite region of bound states from an infinite region of denatured states. The point $(t_0, f = 0)$ with $t_0 = t_m(f = 0)$ corresponds to the traditional melting transition for

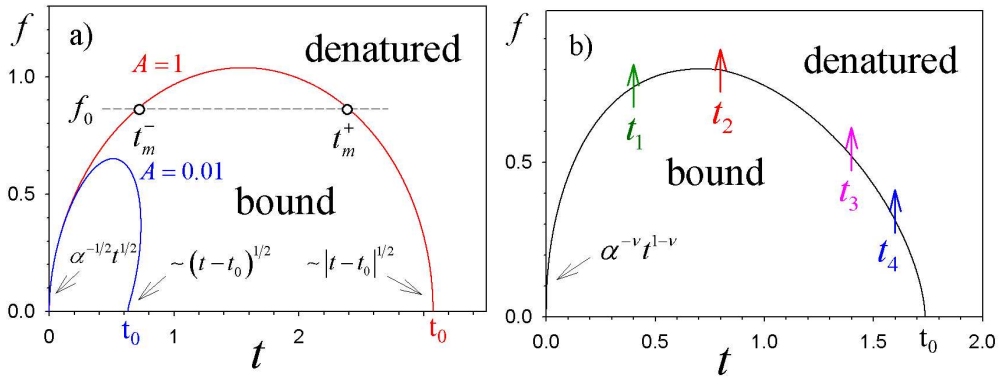

FIG. 4: Transition lines $f_m = F_m a/\varepsilon$ as function of $t = k_B T/\varepsilon$ for $\alpha = 1$, $s = 5$ for denatured loops modelled as (a) ideal random walks and (b) self-avoiding walks. Note the reentrant behaviour at lower temperatures where the required melting force decreases.

free DNA ($F = 0$). The line $f_m(t)$ for $A = 1$ contains a region in which $f_m(t)$ *decreases* with t, such that increased stretching forces f lower the melting temperature $t_m(f)$, corresponding to force-induced destabilisation of DNA [30]. Interestingly, for $A = 0.01$ the line $f_m(t)$ is not single-valued. Moreover, $f_m(t)$ vanishes for both $t \to t_0$ (as $|t - t_0|^{1/2}$) and $t \to 0$ (as $\alpha^{-1/2} t^{1/2}$). This "reentrant behaviour" [36] means that for given $0 < f_0 < f_{\max}$, where $f_{\max}$ is the maximum of $f_m(t)$, the chain does not only denature at a large $t_m^+(f_0)$ but also at a small $t_m^-(f_0)$. This behaviour can be traced back to a balance of the terms $(\beta F a)^2$ and $\beta F a$ in $z_m(F) = \exp(-\alpha y^2)/s$ and Eq. (4), respectively. For $(\beta F a)^2 \ll \beta F a$, i.e., $k_B T \gg F a$, the melting transition at $t_m^+(f_0)$ is mainly driven by the entropy gain on creation of fluctuating loops, similar as for free DNA. For $k_B T \ll F a$ the transition at $t_m^-(f_0)$ is due to the fact that $B[z_m(F), \omega, F]$ *decreases* with $y = \beta F a = f/t$ in the denatured state, due to the rapid decay of $z_m(F)$ [cf. Eq. (4)] [90]. Fig. 4b shows the line $f_m(t)$ for self-avoiding loops with $c = 1.85$ demonstrating analogous behaviour.

At very high forces corrections to this treatment are expected. However the fact that already at moderate (in fact, any positive) external force F the value of the critical exponent c changes indicates that the force-induced denaturation employed in single molecule experiments is physically different from thermal denaturation. This is intuitively clear as the pulling alters not only the free energy of intact base-pairs but also the number of accessible degrees of freedom of the polymer loops forming the denaturation bubbles.

We here treat the DNA denaturation in presence of the external stretching force in analogy

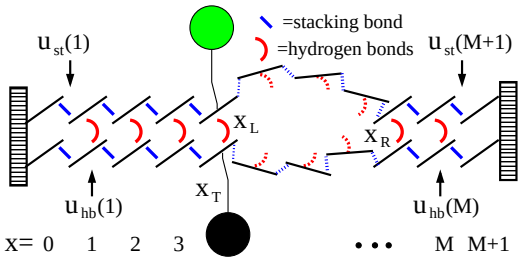

FIG. 5: Clamped DNA domain with internal bps $x = 1$ to M, statistical weights $u_{\rm hb}(x)$, $u_{\rm st}(x)$, and tag position x_T. The DNA sequence enters through the statistical weights $u_{\rm st}(x)$ and $u_{\rm hb}(x)$ for disrupting stacking and hydrogen bonds respectively. The bubble breathing process consists of the initiation of a bubble and the subsequent motion of the forks at positions x_L and x_R. See [25] for details.

to thermal denaturation. The transition, that is, goes from the double-stranded state to fully denatured single-strand. While this view is in accord with a large body of experiments [19, 37–39] and theoretical approaches [30, 40, 41]. one cannot exclude the possibility that an intermediate state of DNA exists, so called S-DNA. A number of recent contributions address this question [41–46] but for now this point remains unresolved.

III. SINGLE DNA BUBBLE DYNAMICS

Below the melting temperature T_m, DNA bubbles are intermittent, i.e., they form spontaneously due to thermal fluctuations and after some time close again. DNA-breathing can be thought of as a biased random walk in the phase space spanned by the bubble size m and its position denoted, e.g., by the left zipper fork position x_L [24, 25]. The bubble creation can be viewed as a nucleation process, whereas the bubble lifetime corresponds to the survival time of the first passage problem of relaxing to the $m = 0$ state after a random walk in the $m > 0$ halfspace [24, 25, 47–49]. Apart from NMR techniques [6, 11] bubble breathing could be measured on the single DNA-bubble level by fluorescence correlation spectroscopy [14]. This technique employs a designed stretch of DNA, in which weaker AT bps form the bubble domain, that is clamped by stronger GC bonds. In the bubble domain, a fluorophore-quencher pair is attached, see Fig. 5. Once the bubble is created, fluorophore and quencher are separated, and fluorescence occurs.

A. Continuum approach for homopolymer DNA

Originally bubble breathing was considered in a random energy model with scaling arguments and numerical solution [50] and for a homopolymer by mapping on a Fokker-Planck equation for a random walker in the bubble free energy landscape with approximate analytical and numerical solution [47]. An analytical approach to bubble breathing in a homopolymer DNA with explicit solution for the distribution of bubble lifetimes is indeed possible by mapping onto the quantum Coulomb problem [51, 52] as we discuss here. In the following subsection we consider explicitly given DNA sequences in a discrete approach.

The Poland-Scheraga free energy for a single bubble has the continuum form [47, 51]

$$\mathscr{F} = \gamma_0 + \gamma x + ck_BT \ln x. \tag{10}$$

in terms of the bubble size $x \geq 0$. Expression (10) corresponds to a logarithmic sink in $\mathscr{F}$ at $x = 0$ and we recognise from this equation that a characteristic bubble size is set by $x_1 = ck_BT/|\gamma|$. We rewrite the free stacking energy in terms of $\gamma \equiv \gamma_1(T_m - T)/T_m$ through the melting temperature T_m, and similarly, we introduce $\epsilon = \gamma_1/[2k_B]\,(T_m^{-1} - T^{-1})$.

For large bubble size $x > x_1$ the linear term dominates and the free energy grows like $\mathscr{F} \sim \gamma_0 + \gamma x$. For small bubbles $x < x_1$ [or close to T_m, where $\gamma(T) \approx 0$] the free energy is characterised by the logarithmic sink but has strictly speaking a minimum at $\mathscr{F} = \gamma_0$ for zero bubble size. We distinguish two temperature ranges: (i) For $\gamma < 0$, i.e., $T > T_m$, $\mathscr{F}$ has a maximum $\mathscr{F}_{\max} = \gamma_0 + ck_BT(\log x_1 - 1)$ at $x = x_1$. The free energy profile thus defines a Kramers escape problem in the sense that an initial bubble can grow in size corresponding to the complete denaturation of the double stranded DNA. The escape probability $P_{\rm esc} \propto \exp(-\Delta\mathscr{F}/k_BT)$, where the free energy barrier is $\Delta\mathscr{F} = ck_BT(\log x_1 - 1)$. Thus

$$P_{\rm esc} \propto \left(\frac{ck_BT}{|\gamma|}\right)^{-c} \tag{11}$$

has a power-law dependence on temperature typical for entropic barriers. In contrast a Kramers escape across a high energetic barrier leads to an Arrhenius behaviour. An example for the latter would be the initiation process of a bubble during which the barrier $\sigma_0 = \exp(-\beta F_s)$ needs to be crossed.

(ii) For $\gamma > 0$, i.e., $T < T_m$, the free energy increases monotonically from $\mathscr{F} = \gamma_0$ at

$x = 0$ and the finite size bubbles are stable. The change of sign of γ at $T = T_m$ thus defines the bubble melting.

The gradient of the free energy profile then enters as force term in a Langevin equation for the bubble size x. Such a treatment is possible since x is the slow variable of the system compared to the polymeric degrees of freedom of a bubble and even the entire chain unless the chain size becomes too large. The Langevin equation can then be mapped onto the Fokker-Planck equation for the probability density $P(x,t)$ to find a bubble of size x at time t:

$$\frac{\partial P(x,t)}{\partial t} = D k_B T \frac{\partial}{\partial x}\left(\left[\frac{c}{x} - \epsilon\right] + \frac{\partial^2}{\partial x^2}\right) P(x,t). \tag{12}$$

Here D is the noise strength of the thermal environment measured in units of $k_B T$ and time. It is now the task to derive from this dynamical description physically relevant and measurable quantities. These are the bubble lifetime and its distribution as well as autocorrelation functions of the bubble dynamics. We here concentrate on the former while addressing the autocorrelation function in the subsequent section dealing with the discrete formalism. More details on the autocorrelation function in the continuum limit can be found in Ref. [51–53].

The single bubble dynamics can be analysed in different ways; namely in terms of the underlying Langevin equation including the interpretation of the single bubble dynamics below the melting temperature as a noisy finite time singularity. Alternatively a weak noise analysis allowing one to interpret the dynamics through orbitals in phase space portraits. Finally, one may turn to the Fokker-Planck equation (12). For more details we refer to Refs. [47, 51, 52].

To determine the lifetime distribution of a bubble once opened we face a technical problem posed by the c/x term in the drift term of Eq. (12). One way to circumvent this is to map this Fokker-Planck equation onto the corresponding imaginary time Schrödinger equation of the quantum Coulomb problem [51, 52]. From this formulation one is able to deduce the behaviour of the bubble lifetime. We distinguish three cases.

(i) Below the melting temperature. At $T < T_m$ one can determine the density of the bubble lifetime distribution analytically in the long time limit obtaining

$$\wp(t) \simeq x_0^{1+c} e^{|\epsilon| x_0} e^{-\epsilon^2 t/2} t^{-3/2-c/2} \tag{13}$$

Thus, we observe a power-law behaviour $t^{-3/2-c/2}$ with an exponential cutoff at $\tau = 2/\epsilon^2$ such that the bubble lifetime is always finite. This form for $\wp(t)$ generalises the expression

of the first passage time density of a bubble without entropy loss correction (i.e., $c = 0$) with constant drift $|\epsilon|$ towards bubble closure [47]. For the mean bubble lifetime we find the approximate expression

$$\mathcal{T} = \int_0^\infty t\wp(t)dt \simeq \frac{x_0}{|\epsilon|}\frac{K_{(c-1)/2}(x_0|\epsilon|)}{K_{(c+1)/2}(x_0|\epsilon|)}. \tag{14}$$

For large sufficiently large values of $x_0|\epsilon|$ the ratio of the two Bessel functions tends to 1, in particular, for the Gaussian chain limit $c = 3/2$ we find $K_{1/2}(x_0|\epsilon|)/K_{3/2}(x_0|\epsilon|) = 1/(1 + |\epsilon|/x_0)$. This result for $\mathcal{T}$ includes the characteristic bubble lifetime $x_0/|\epsilon|$ when the loop entropy correction is neglected ($c = 0$) [47].

(ii) At the melting temperature. Right at $T = T_m$ the drift exerted by the free stacking energy ϵ vanishes, and the dynamics is almost free diffusion. The result for the density of bubble lifetimes reads

$$\wp(t) = \frac{2x_0^{1+c}}{\Gamma(1/2 + c/2)}e^{-x_0^2/2t}(2t)^{-3/2-c/2} \tag{15}$$

and is normalised and exact for all times. In this case the power-law $t^{-3/2-c/2}$ determines the long time behaviour. While for free diffusion ($c = 0$) the corresponding mean bubble lifetime $\int_0^\infty t\wp(t)dt$ diverges [47], for all $c > 1$ we encounter the mean bubble lifetime

$$\mathcal{T} = \frac{x_0^2}{c-1} \tag{16}$$

which interestingly grows like the square of the initial bubble size in contrast to the linear scaling in the case of diffusion with linear drift in case (i). In addition to the finite mean bubble lifetime a value $c > 2$ would also cause a power-law decay $C(t) \sim t^{1-c/2}$ of the associated correlation function at long times in contrast to the plateau $C(t) \sim 1$ reached for $1 < c < 2$ [53].

(iii) Above the melting temperature. At $T > T_m$ the situation is opposite to case (i); namely the drift is now directed towards the complete denaturation of the chain. In a long chain the one bubble picture would no longer hold and bubble coalescence needs to be taken into account. However in shorter DNA constructs preferring one single bubble the density of bubble lifetimes would decay exponentially [51, 52].

B. Discrete approach and sequence dependence

The natural coordinate for the unzipping and zipping of base pairs in DNA breathing dynamics is the location x of a respective base pair along the chemical backbone of the DNA

molecule. By its very nature this is a discrete variable. While in the continuum approach one may include certain given distributions of more and less stable regions (predominantly GC-rich versus predominantly AT-rich) the use of a truly discrete x allows one to consider any given sequence. This is of particular importance when analysing actual biologically relevant sequences or those designed sequences that are used in a given experiment. Such a discrete approach in terms of the master equation will be described here. We note that a disadvantage of this method is the limited system size one can de facto analyse due to computational constraints.

With a discrete coordinate we are also able to explicitly distinguish hydrogen bonding and stacking energies and use the parameters for the free energies from Krueger et al. [6]. For the setup sketched in Fig. 5 we then find the partition function. The positions x_L and x_R of the zipper forks correspond to the right- and leftmost closed bp of the bubble. x_L and x_R are stochastic variables, whose time evolution in the energy landscape defined by the partition factor ($m \geq 1$)

$$\mathscr{Z}(x_L, m) = \frac{\xi'}{(1+m)^c} \prod_{x=x_L+1}^{x_L+m} u_{\mathrm{hb}}(x) \prod_{x=x_L+1}^{x_L+m+1} u_{\mathrm{st}}(x) \tag{17}$$

characterises the bubble dynamics. $\mathscr{Z}$ is written in terms of x_L and bubble size $m = x_R - x_L - 1$, with $\mathscr{Z}(m=0) = 1$. Here, $\xi' = 2^c\xi$, where $\xi \approx 10^{-3}$ is the ring factor for bubble initiation from Ref. [6] that is related to the cooperativity parameter $\sigma_0 \approx 10^{-5}$ [7, 54] by $\sigma_0 = \xi \exp(\epsilon_{\mathrm{st}})$ [6]. For the entropy loss on forming a closed polymer loop we assign the factor $(1+m)^{-c}$ [54, 55] and take $c = 1.76$ for the critical exponent [33]. This corresponds to the Flory form 3ν for the entropy loss factor for a polymer ring with excluding volume. The best known value for ν is 0.588 [56–58]. Note that there exist alternative models taking into account the self-avoiding interactions of the bubble with the rest of the chain, leading to an increased value for c ($c \approx 2.1$) such that the denaturation transition becomes first order [31, 33]. Note also that a bubble with m open bps requires breaking of m hydrogen bonds and $m+1$ stacking interactions.

The zipper forks move stepwise $x_{L/R} \to x_{L/R} \pm 1$ with rates $\mathsf{t}^{\pm}_{L/R}(x_{L/R}, m)$. We define for bubble size decrease

$$\mathsf{t}^{+}_{L}(x_L, m) = \mathsf{t}^{-}_{R}(x_L, m) = k/2 \qquad (m \geq 2) \tag{18}$$

for the two forks [91]. The rate k characterises a single bp zipping. Its independence of x corresponds to the view that bp closure requires the diffusional encounter of the two bases

and bond formation; as sterically AT and GC bps are very similar, k should not significantly vary with bp stacking. k is the only adjustable parameter of our model, and has to be determined from experiment or future MD simulations. The factor $1/2$ is introduced for consistency [48, 49]. Bubble size increase is controlled by

$$\begin{aligned} \mathsf{t}_L^-(x_L,m) &= k u_{\rm st}(x_L) u_{\rm hb}(x_L) s(m)/2, \\ \mathsf{t}_R^+(x_L,m) &= k u_{\rm st}(x_R+1) u_{\rm hb}(x_R) s(m)/2, \end{aligned} \tag{19}$$

for $m \geq 1$, where $s(m) = \{(1+m)/(2+m)\}^c$. Finally, bubble initiation and annihilation from and to the zero-bubble ground state, $m = 0 \leftrightarrow 1$ occur with rates

$$\begin{aligned} \mathsf{t}_G^+(x_L) &= k\xi' s(0) u_{\rm st}(x_L+1) u_{\rm hb}(x_L+1) u_{\rm st}(x_L+2) \\ \mathsf{t}_G^-(x_L) &= k. \end{aligned} \tag{20}$$

The rates t fulfil detailed balance conditions. The annihilation rate $\mathsf{t}_G^-(x_L)$ is twice the zipping rate of a single fork, since the last open bp can close either from the left or right. Due to the clamping, $x_L \geq 0$ and $x_R \leq M+1$, ensured by reflecting conditions $\mathsf{t}_L^-(0,m) = \mathsf{t}_R^+(x_L, M-x_L) = 0$. The rates t together with the boundary conditions fully determine the bubble dynamics.

In the FCS experiment fluorescence occurs if the bps in a Δ-neighbourhood of the fluorophore position x_T are open [14]. Measured fluorescence time series thus correspond to the stochastic variable $I(t)$, that takes the value 1 if at least all bps in $[x_T - \Delta, x_T + \Delta]$ are open, else it is 0. The time averaged ($\overline{\cdot}$) fluorescence autocorrelation

$$A_t(x_T, t) = \overline{I(t)I(0)} - \overline{I(t)}^2 \tag{21}$$

for the sequence AT9 from [14] are rescaled in Fig. 6.

DNA breathing is described by the probability distribution $P(x_L, m, t)$ to find a bubble of size m located at x_L whose time evolution follows the master equation

$$\frac{\partial P(x_L,m,t)}{\partial t} = \mathbb{W} P(x_L,m,t). \tag{22}$$

The transfer matrix $\mathbb{W}$ incorporates the rates t. Detailed balance guarantees equilibration toward

$$P_{\rm eq} = \lim_{t\to\infty} P(x_L,m,t) = \frac{\mathscr{Z}(x_L,m)}{\mathscr{Z}}, \tag{23}$$

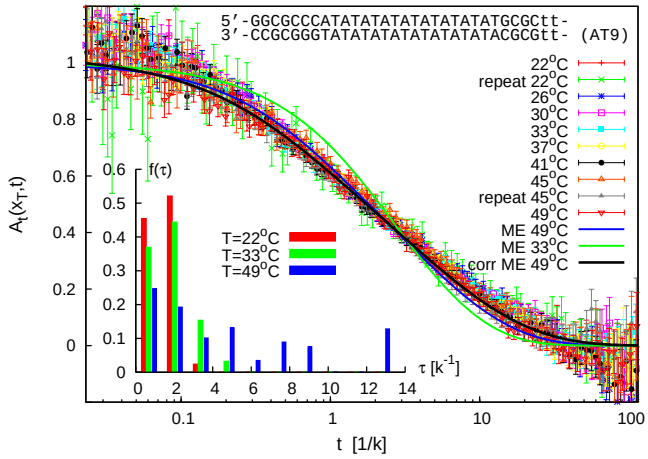

FIG. 6: Scaling plot of $A_t(x_T, t)$ at various T for the sequence AT9 from [14] as indicated in the figure. This experimental construct is designed with a weak AT-rich bubble domain in the core, a GC clamp at both ends and additional bulge loop of DNA single strand consisting of four T bases. The symbols represent experimental data at various temperatures, see Refs. [24, 25] for more details. We also include results from our master equation model. Inset: Relaxation time spectrum. See text for more details.

with $\mathscr{Z} = \sum_{x_L,m} \mathscr{Z}(x_L, m)$ [48, 49, 59]. The master equation and the explicit construction of $\mathbb{W}$ are discussed at length in Refs. [25, 48, 49, 60]. Eigenmode analysis and matrix diagonalisation produces all quantities of interest such as the ensemble averaged autocorrelation function

$$A(x_T, t) = \langle I(t)I(0)\rangle - (\langle I\rangle)^2. \tag{24}$$

$\langle I(t)I(0)\rangle$ is proportional to the survival density that the bp is open at t and that it was open initially [24, 60].

In Fig. 6 the blue curve shows the predicted behaviour of $A(x_T, t)$, calculated for $T = 49°\mathrm{C}$ with the parameters from [6]. As in the experiment we assumed that fluorophore and quencher attach to bps x_T and $x_T + 1$, that both are required open to produce a fluorescence signal. From the scaling plot, we calibrate the zipping rate as $k = 7.1 \times 10^4/\mathrm{s}$, in good agreement with the findings from Ref. [14]. The calculated behaviour reproduces the data within the error bars, while the model prediction at $T = 35°\mathrm{C}$ shows more pronounced deviation. Potential causes are destabilising effects of the fluorophore and quencher, and additional modes that broaden the decay of the autocorrelation. The latter is underlined

by the fact that for lower temperatures the relaxation time distribution $f(\tau)$, defined by $A(x_T,t) = \int \exp(-t/\tau) f(\tau) d\tau$, becomes narrower (Fig. 6 inset). Deviations may also be associated with the correction for diffusional motion of the DNA construct, measured without quencher and neglecting contributions from internal dynamics [61]. Indeed, the black curve shown in Fig. 6 was obtained by a 3% reduction of the diffusion time [92] which should roughly account for the presence of the quencher.

Stochastic simulation. Based on the rates t, stochastic simulations give access to single bubble fluctuations [62]. The corresponding Gillespie algorithm uses the joint probability density of waiting time τ and path $\mu = +/-$,

$$P(\tau,\mu,\nu) = \mathsf{t}^{\mu}_{\nu}(x_L,m) \exp\left(-\tau \sum_{\mu,\nu} \mathsf{t}^{\mu}_{\nu}(x_L,m)\right), \tag{25}$$

defining for given state (x_L,m) after what time τ the next step of fork $\nu \in \{L,R\}$ occurs. The formulation via the waiting time density $\sum_{\mu,\nu} P$ is economical computationally, avoiding a large number of unsuccessful opening attempts in traditional Langevin simulations. Using (25) we obtain the single bubble time series in Fig. 7 for two different tag positions in the T7 bacteriovirus promoter sequence

```
      1                   20
      |                   |
5'-ATGACCAGTTGAAGGACTGGAAGTAATACGACTC
   AGTATAGGGACAATGCTTAAGGTCGCTCTCTAGGAG-3'
      |  |                          |
      38 41                        68
```
(26)

whose TATA motif is underlined [23]. A promoter is a sequence (often containing the so called TATA motif) placed at the start of a gene, to which RNA polymerase is then recruited to initiate transcription [63]. Motives such as TATA are believed to assist polymerase during the transcription initiation [22, 25]. Fig. 7 shows the signal $I(t)$ at 37°C for the tag positions $x_T = 38$ in the core of TATA, and $x_T = 41$ at the second GC bp after TATA. Bubble events occur much more frequently in TATA (the TA/AT stacking interaction is particularly weak [6]). This is quantified by the density of waiting times $\psi(\tau)$ spent in the $I = 0$ state, whose characteristic time scale $\tau' = \int_0^\infty d\tau \tau \psi(\tau)$ is more than an order of magnitude longer than at $x_T = 41$. In contrast, we observe similar behaviour for the density of opening times $\phi(\tau)$ for $x_T = 38$ and 41. The solid lines are the results from the master equation showing

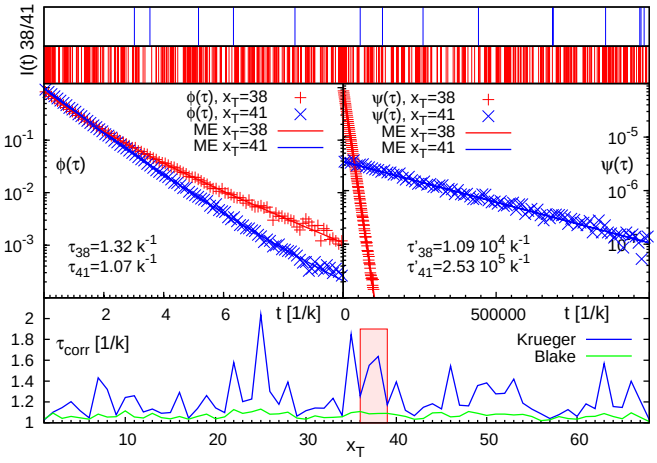

FIG. 7: Top: Time series $I(t)$ for the T7 promoter, for the opening of base pairs at labels $x_T = 38$ (in the TATA motif) and 41 (in the adjacent GC region). Middle: Fluorescence time $\phi(\tau)$ corresponding to the bubble life time and waiting time $\psi(\tau)$ elapsing between bubble events. While the bubble lifetime in both regions of the sequence are approximately equivalent, the occurrence frequency of bubbles is indeed significantly higher within the TATA domain. Bottom: Mean fluorescence time for $\Delta = 0$ for parameter sets from Blake et al. [54] and Krueger et al. [6]. One recognises the much stronger sequence sensitivity for the parameters from Krueger et al. The shaded area corresponds to the TATA domain. Again the lifetime does not appear to significantly distinguish the TATA domain. In contrast the simultaneous opening of 4 sequential base pairs clearly favours opening of the motif [24, 25].

excellent agreement with the results from the Gillespie stochastic simulation. Notice that whereas $\psi(t)$ is characterised by a single exponential, $\phi(t)$ show a crossover between different regimes. For long times both $\psi(\tau)$ and $\phi(\tau)$ decay exponentially as it should for a finite DNA stretch.

C. Bubbles in biological sequences

After presenting our results for the T7 promoter sequence above in this section we comment on the biological relevance of the distribution of soft and hard zones, in particular with respect to transcription initiation. A more detailed analysis can be found in [22–25].

Let us start by briefly commenting on the biochemical relevance of the TATA box motif

(also referred to as Goldberg-Hogness box). It is a DNA sequence (cis-regulatory element) found in the promoter region of most genes in eukaryotes and a group of single-celled microorganisms called archaea. Similar binding motifs with similar properties exist in other organisms. The TATA box is the binding site of transcription factors and is involved in the process of transcription by RNA polymerase. Its core sequence is `5'-TATAAA-3'` or a variant, usually followed by three or more adenine bases. Commonly it is located 25 base pairs upstream to the transcription site. The TATA box is normally bound by the TATA Binding Protein (TBP) during transcription. The TBP unwinds the DNA and strongly bends it. At a later stage the TATA box is bound by RNA polymerase and transcription commences.[93] The high proneness towards bubble formation at the TATA box is therefore believed to actively contribute to transcription initiation.[94]

1. Bacteriophage T7 core promoter.

Its sequence is displayed in Eq. (26). It contains the TATA box at base pair labels 36 to 39. Fig. 8 shows the equilibrium probabilities for the base pairs to be open. In this example the TATA box is located right next to the transcription start site. From the graph one can see that indeed the simultaneous opening probability of four base pairs is significantly increased at the position of the TATA box. Note the level of the opening probability of a random sequence also drawn in the figure. Accordingly several domains of significantly increased bubble probability exist along this sequence.

2. Adenovirus Major Late Promoter.

Its 86 base pair sequence

```
      1                              31
      |                              |
5'- GCCACGTGACCAGGGGTCCCCGCCGGGGGGGT
    ATAAAAGGGGCGGACCTCTGTTCGTCCTC                    (27)
    ACTGTCTTCCGGATCGCTGTCCAG -3'
    |                    |
   TSS                   84
```

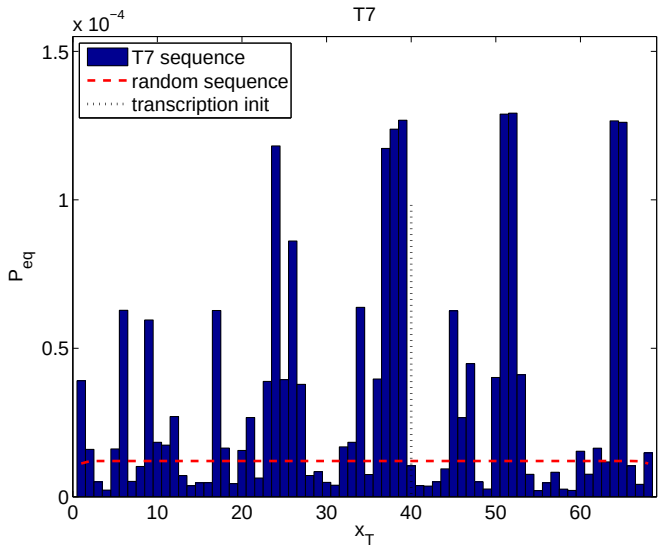

FIG. 8: Equilibrium opening probability of base pairs in the sequence of the bacteriophage T7 core promoter.

contains a transcription start site at the position labelled TSS, compare Fig. 9. In this example the TATA box is located upstream at the base pair label -29. In this example the TATA box is extremely more likely to simultaneously open than any other domain along the sequence.

3. Adeno Associated Viral P5 promoter.

This sequence consists of the 69 base pairs

```
          1                        25
          |                        |
5'- GTGGCCATTTAGGGTATATATGGCCGAGTGAGCGA
    GCAGGATCTCCATTTTGACCGCGAAATTTGAACG -3'
              |                        |
             TSS                      67
```
(28)

and supports binding of TBP at the TATA box as well as the binding of the Yin Yang 1 (YY1) transcription factor. YY1 is known to interact with the TBP [64]. YY1 binds to a specific sequence element of the form `CCATNTT` marked blue in the sequence. As can be seen from Fig. 10 these two binding motifs have a significantly higher cooperative opening

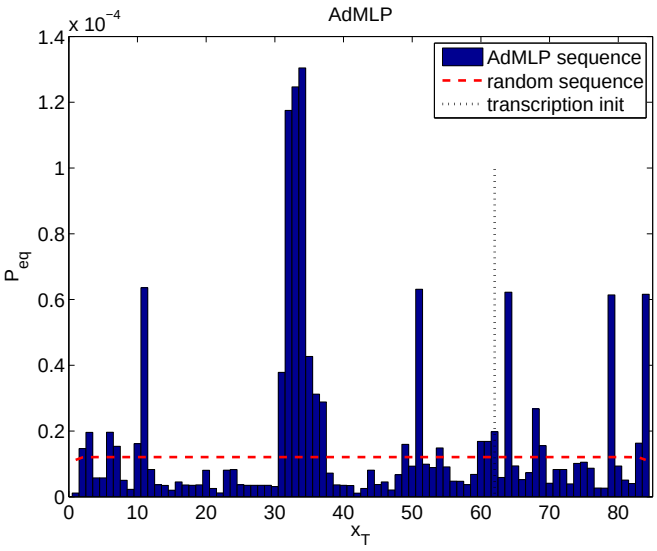

FIG. 9: Equilibrium opening probability of base pairs in the sequence of the Adenovirus Major Late Promoter.

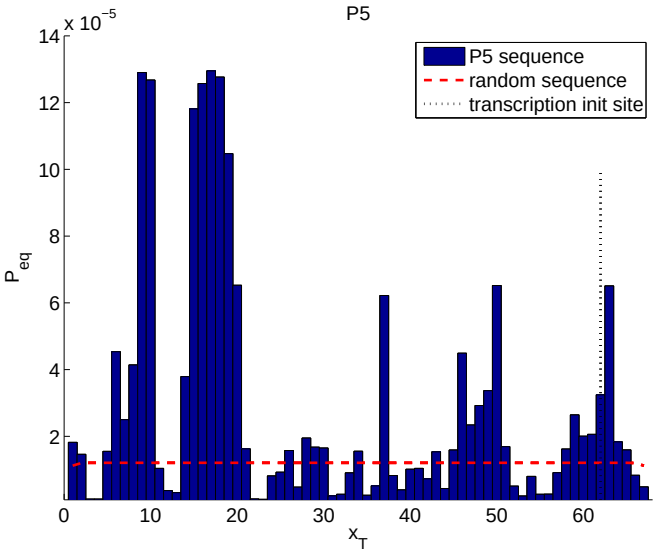

FIG. 10: Equilibrium opening probability of base pairs in the sequence of the Adeno Associated Viral P5 promoter.

probability than any other sequence element of this promoter. The analysis also shows a broader but lower peak around the transcription start site.

In summary this analysis shows that indeed local instability of the DNA sequence ap-

pears to occur at specific binding sequences for proteins involved in transcription initiation. Whether it is just the lower free energy needed to break these sequences or indeed rare bubble openings at these site that help the protein binding remains an open question.

IV. DNA BUBBLE COALESCENCE

It has been shown in a quantitative analysis that the experimentally accessible autocorrelation function is sensitive to the stacking parameters of DNA [24, 25]. However, it has not been fully appreciated to what extent the fluorophore and quencher molecules, that are attached to the DNA construct in the experiments reported in Refs. [14, 61, 65], influence the stability of DNA. Moreover, the zipping rates measured in the single molecule fluorescence setup differ from those determined in NMR experiments [11, 14]. We here propose and study an alternative setup for the single molecule fluorescence investigation of DNA breathing as shown in Fig. 11 that may improve and complement the single molecule data obtained from a DNA construct with a single bubble domain.[95] In this setup, a short stretch of DNA, clamped at both ends, is designed such that two soft zones consisting of weaker AT-bps are separated by a more stable barrier region rich in GC bps. For simplicity, we assume that both soft zones and barrier are homopolymers with a bp-dissociation free energy ΔG_s and ΔG_b, respectively, and, in accordance with the experimental findings of reference [14], we neglect secondary structure formation in the barrier zone. At temperatures higher than the melting temperature T_s of the soft zones but still lower than the melting temperature T_b of the barrier region, thermal fluctuations will gradually dissociate the barrier, until the two bubbles coalesce. Once coalesced, the release of the free energy corresponding to one cooperativity factor $\sigma_0 \approx 10^{-5\ldots-3}$ is released, stabilising the coalesced bubble against reclosure of the barrier. Moreover there exists a significant dynamic barrier stemming from the necessity of diffusional encounter of the two bases in order to reanneal the barrier. Both points lead to a long lifetime of the coalesced state. This fact should allow for a meaningful measurement of the coalescence time in experiment, and therefore provide a new and sensitive method to measure DNA stability data and base pair zipping rates. We also study the case when the system is prepared as above and then T suddenly increased such that $T > T_b > T_s$ so that the system is driven towards coalescence. In both cases the two boundaries between bubbles and barrier perform a (biased) random walk in opposite free energy potentials.

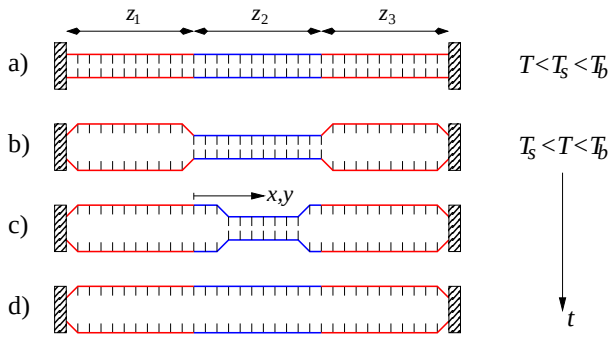

FIG. 11: Schematic of the DNA construct for bubble coalescence. Note that the position of both ends of the barrier region are measured from the same point (the position of the leftmost barrier base pair).

The statistical weight of the construct before coalescence,

$$\mathscr{Z}_{X,Y} = (\xi e^{N_L \beta\varepsilon'}) e^{(X-Y+N)\beta\varepsilon} (\xi e^{N_R \beta\varepsilon'}), \tag{29}$$

at $T_b > T > T_s$ involves the cooperativity factor $\xi \approx 10^{-5}$ for each bubble, and a Boltzmann factor for each broken bp with free energies $\varepsilon' > 0$ and $\varepsilon < 0$, compare reference [24]. Upon coalescence, the boundary free energy corresponding to one factor ξ is released,

$$\mathscr{Z}_{\rm coal} = \xi e^{(N_L+N_R)\beta\varepsilon' + N\beta\varepsilon}, \tag{30}$$

stabilising the system against immediate transition back to a two-bubble state. It is this distinctive feature that should render this setup an interesting model system for single molecule analyses of DNA denaturation dynamics as the coalesced state can be determined by measuring first passage time statistics (corresponding to the introduction of an absorbing boundary condition at the point of coalescence).

In our analysis we use a continuum approach to the stochastic motion of the two zipping forks at either end of the barrier zone with locations x and y. The probability density $P(x, y, t)$ then follows the bivariate Fokker-Planck equation [66]

$$\frac{\partial}{\partial t} P(x, y, t) = \left(\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right] - 2f \frac{\partial}{\partial x} + 2f \frac{\partial}{\partial y} \right) P(x, y, t), \tag{31}$$

with the dimensionless force $f = N(u-1)/(1+u)$ and time rescaled by $k(1+u)/2N^2$. Equation (31) is completed by the initial condition $P(x, y, 0) = \delta(x - x_0)\delta(y - y_0)$ and the

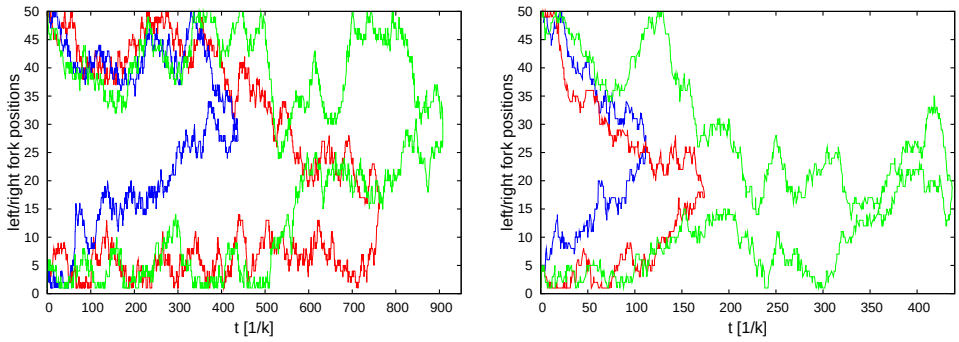

FIG. 12: Trajectories of the random motion of the two bubble forks.

reflecting boundary conditions (the bubbles in the soft zones are assumed to be open at all times)

$$\left((\frac{\partial}{\partial}x - 2f\right) P(x,y,t)\Big|_{x=0} = \left(\frac{\partial}{\partial y} + 2f\right) P(x,y,t)\Big|_{y=1} = 0. \tag{32}$$

Moreover, we impose the absorbing boundary condition $P(x,x,t) = 0$. This defines the vicious walker property [67], terminating the process when the two walkers meet. The fact that the two walker move in opposite potentials actually make this problem a previously unsolved case of vicious walkers models [66].

Typical examples of individual trajectories resulting from a Gillespie algorithm are displayed in figure 12, where traces of the two interfaces (forks) cornering the barrier region are shown. Bubble coalescence terminates each pair of trajectories.

The analysis in Ref. [66] reveals the distribution of coalescence positions (*i.e.*, where the two zipper forks eventually meet) and the coalescence times, as shown in Fig. 13. The curves for the PDF $\rho(x)$ of the coalescence position exhibit a pronounced crossover from a relatively sharply peaked form to an almost flat behaviour. The former occurs for large positive force f, corresponding to a strong drift toward a potential well, with negligible influence of the boundary conditions. In contrast, for large negative f, corresponding to a high barrier for coalescence, the insensitivity of $\rho(x)$ to the position x can be explained in terms of a simple Arrhenius argument: The probability of the walker to be at a position x is proportional to the Boltzmann weight, $\exp(-\beta\phi(x))$, where $\phi(x) = -\int^x F(x')dx'$ is the free energy corresponding to the force $F(x)$. Then, the joint probability to have both walkers meet at the same position is given by the product $\exp(-\beta[\phi_L(x)+\phi_R(x)]) \approx const.$ as the two walkers are in opposite linear potentials and the position dependence of the exponent cancels out. This

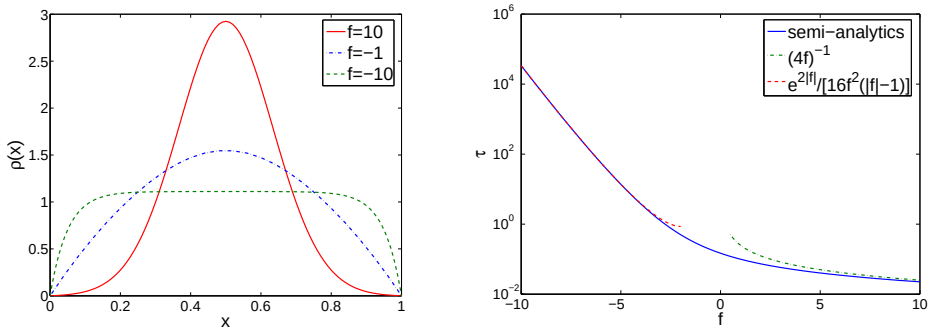

FIG. 13: Left: Distribution of coalescence positions within the rescaled barrier zone $[0, 1]$. Right: Distribution of coalescence times.

simple picture necessarily breaks down close to the boundaries. (ii) The f-dependence of the mean first passage time τ crosses over from the $\tau \simeq 1/f$ behaviour typical for diffusion in a strong positive force pushing the two walkers together, to the exponential form $\tau \simeq \exp(2|f|)$ of the associated Kramers problem. The former problem was studied in reference [47] by neglecting the boundaries and switching to the relative coordinate description which enables one to find the analytic result $\tau = 1/(4f)$. For the Kramers problem ($f \ll -1$) the analytic solution for both $\rho(x) = [1 - e^{-2|f|x} - e^{-2|f|(1-x)}]|f|/(|f| - 1)$ and $\tau = e^{2|f|}/[16f^2(|f| - 1)]$ can be found rather easily [68] by the expansion into the lowest two eigenmodes of $p(x, t|x_0)$.

V. COUPLED DYNAMICS OF DNA BUBBLES AND SELECTIVELY SINGLE-STRAND DNA BINDING PROTEINS

A traditional puzzle had been the question why the presence of selectively single-strand DNA binding proteins (SSBs) does not lead to full denaturation of the DNA [1]. While ideas about a kinetic block were brought forth relatively early [69–71], experimentally this puzzle could only be solved by single molecule methods in which the denaturation was not induced by temperature but force. In a series of experiments the binding and unbinding kinetics of SSBs and their mutants to DNA denaturation bubbles and the resulting effect on the denaturation force were studied in great detail [19, 72, 73]. Here we discuss a simple model for the SSB-DNA interaction in a homopolymer approach by a master equation approach [48, 49].

The quantity of interest is the joint probability $P(m, n, t)$ to have a bubble consisting of

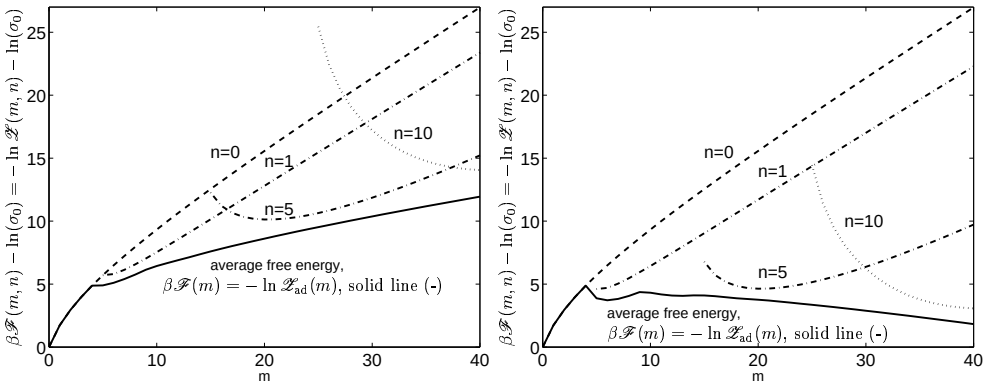

FIG. 14: Effective free energy of the SSB-DNA bubble interaction in the limit $\gamma \gg 1$ (—), and free energy landscape for various fixed n ($u = 0.6$, $M = 40$, $c = 1.76$, $\lambda = 5$). Left: $\kappa = 0.5$; Right: stronger binding, $\kappa = 1.5$. In the latter case the binding strength of the SSB suffices to cause a decreasing effective free energy and therefore induce full denaturation of the DNA. Due to the finite size effects the nucleation barrier for initiation of SSB exchange has to be crossed.

m broken bps, and n SSBs bound to the two arches of the bubble. In addition to the rates for bubble increase and decrease, the rates for SSB binding and unbinding are necessary to define the breathing dynamics in the presence of SSBs. On the statistical level, the effect of the SSBs becomes coupled to the motion of the zipper forks. Thus, the rate for bubble size decrease is proportional to the probability that no SSB is located right next to the corresponding zipper fork; and the rate for SSB binding is proportional to the probability that there is sufficient unoccupied space on the bubble. Binding is allowed to be asymmetric, and is related to a parking lot problem in the following sense. The number λ of bases occupied by a bound SSB is usually (considerably) larger than one. In order to be able to bind in between two already bound SSBs, the distance between these two SSBs must be larger than λ. The larger λ the less efficient the SSB-binding becomes, similar to parking large cars on a parking lot designed for small vehicles. Apart from the binding size λ of the SSBs, two additional physical parameters come into play: the unbinding rate q of the SSBs, and their binding strength $\kappa = c_0 K^{\rm eq}$ consisting of the volume concentration c_0 of SSBs and the equilibrium binding constant $K^{\rm eq} = v_0 \exp(\beta |E_{\rm SSB}|)$, with the typical SSB volume v_0 and binding energy $E_{\rm SSB}$.

The coupled dynamics of SSB-binding and bubble breathing is discussed in references [48, 49]; similar effects in end-denaturing DNA were studied in [65] in detail. Here, we

report the behaviour of the effective free energy landscape in the limit of fast SSB-binding in the sense that the dimensionless parameter $\gamma \equiv q/k$ of SSB-unbinding and bubble zipping rates is large, $\gamma \gg 1$. This limit allows one to average out the SSB-dynamics and to calculate an effective free energy, in which the bubble dynamics with the slow variable m runs off. The result for two different binding strengths κ is shown in figure 14, along with the free energies corresponding to keeping n fixed. It is distinct that while for lower κ the presence of SSBs diminishes the slope of the effective free energy, for larger κ the slope actually becomes negative. In the first case, that is, the bubble opening is more likely, but still globally unfavourable. In the latter case, the presence of SSBs indeed leads to full denaturation. One observes distinct finite size effects due to $\lambda > 1$: only when the bubble reaches a minimal size $m \geq \lambda$, SSB-binding may occur, a second SSB is allowed to bind to the same arch only once $m \geq 2\lambda$, etc. This effect also produces the nucleation barrier for full denaturation in the right plot of figure 14. Similar finite size effects were investigated for biopolymer translocation in references [74, 75]. We note that the transition to denaturation could also be achieved by reaching a smaller positive slope of the effective free energy in the presence of SSBs, and additional titration or change of the effective temperature through actual temperature change or mechanical stretching as performed in the experiments reported in references [19, 72, 73].

VI. CONCLUDING REMARKS

DNA possesses a number of properties that render it a very attractive model system. Thus the study of the DNA denaturation transition has occupied statistical physics for around five decades. DNA is comparatively thin and stiff locally while its overall length is fully macroscopic. Thus it is probably the closest available example for testing the predictions from polymer physics. In particular single DNA can be probed and manipulated and its interactions with binding proteins and chemicals investigated. This includes the monitoring of single DNA bubbles and their interaction with specifically single stranded DNA binding proteins, both described here, as well as the interaction of DNA with intercalators [37]. By now single molecule assays can also be used to study the search mechanisms of DNA binding proteins scanning it for specific binding sites relevant in gene regulation and DNA repair [20, 76–78]. This attractiveness of DNA combines with its ultimate rôle as the molecule of

life and therefore is one of the finest examples where the interests of biological physics meet those of biochemists and molecular biologists.

The label *century of biology* is frequently bestowed upon the 21st (e.g., Ref. [79]). In the wash of the success of biology, molecular and systems biology in particular, one experiences a mushrooming number of works in the biological physics sector. Indeed many of these problems pose very attractive and new questions to physicists and along with the availability of single molecule techniques prompt new advances, for instance in statistical physics.

A prime example for the challenges ahead is the current lack of understanding of biochemical processes in living cells under conditions of molecular crowding [80–82]. It is being realised that knowledge obtained under dilute conditions *in vitro* does not necessarily translate to the situation *in vivo* and this point will need considerable more quantitative investigation. As it stands the input from biological physics will be crucial, for example regarding diffusive processes. It appears that subdiffusion of biopolymers occurs in conditions of molecular crowding [83–85] this being the likely source for strong scatter of time averages and apparent diffusivities of single trajectories [83, 86, 87] requiring great care in the quantitative analysis [88].

---

[1] A. Kornberg and T. A. Baker, *DNA Replication* (W. H. Freeman, New York, 1992).

[2] J. D. Watson and F. H. C. Crick, Nature **171**, 737 (1953).

[3] M. D. Frank-Kamenetskii, Phys. Rep. **288**, 13 (1997).

[4] S. G. Delcourt and R. D. Blake, J. Biol. Chem. **266**, 15160 (1991).

[5] J. SantaLucia Jr., Proc. Natl. Acad. Sci. **95**, 1460 (1998).

[6] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, Biophys. J. **90**, 3091 (2006).

[7] D. Poland and H. A. Scheraga *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).

[8] R. M. Wartell and A. S. Benight Phys. Rep. **126**, 67 (1985).

[9] E. Carlon, M. L. Malki, and R. Blossey, Phys. Rev. Lett. **94**, 178101 (2005).

[10] E. Yeramian, Gene **255**, 139 (2000); *ibid.* 151 (2000).

[11] M. Guéron, M. Kochoyan, and J. L. Leroy, Nature **328** 89 (1987).

[12] W. Thumm, A. Seidl, and H.-J. Hinz, Nucl. Acids Res. **16**, 11737 (1988).

[13] P. López-García and P. Forterre, Molec. Microbiol. **23**, 1267 (1997).

[14] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).

[15] S. B. Smith, Y. J. Cui, and C. Bustamante, Science **271**, 795 (1996).

[16] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, Biophys. J. **82**, 3160 (2002).

[17] H. Clausen-Schumann, M. Rief, C. Tolksdorf, and H. E. Gaub, Biophys. J. **78**, 1997 (2000).

[18] R. Metzler, T. Ambjörnsson, A. Hanke, and S. Levene, J. Comp. Theor. Nanoscience **4**, 1 (2007).

[19] K. Pant, R. L. Karpel, and M. C. Williams, J. Mol. Biol. **327**, 571 (2003).

[20] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).

[21] T. Ambjörnsson and R. Metzler, J. Phys. Cond. Mat. **17**, S1841 (2005); Phys. Rev. E **72**, 030901(R) (2005).

[22] C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop, and A. Usheva, Nucl. Acids. Res. **32**, 1584 (2004).

[23] G. Kalosakas, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, Europhys. Lett. **68**, 127 (2004).

[24] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Phys. Rev. Lett. **97**, 128105 (2006).

[25] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Biophys. J. **92**, 2674 (2007).

[26] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).

[27] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **44**, R44 (1993).

[28] B. S. Alexandrov, L. T. Wille, K. Ø. Rasmussen, A. R. Bishop, and K. B. Blagoev, Phys. Rev. E **74**, 050901(R) (2006).

[29] A. Hanke, Martha G. Ochoa, and R. Metzler, Phys. Rev. Lett. **100**, 018106 (2008).

[30] I. Rouzina and V. A. Bloomfield, Biophys. J. **80**, 882; *ibid.* 894.

[31] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000); Eur. Phys. J. B **27**, 132 (2002).

[32] M. E. Fisher, J. Chem. Phys. **44**, 616 (1966).

[33] C. Richard and A. J. Guttmann, J. Stat. Phys. **115**, 925 (2004).

[34] E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002).

[35] P. G. de Gennes, *Scaling concepts in polymer physics* (Cornell University Press, Ithaca, New York, 1979).

[36] E. Orlandini, S. M. Bhattacharjee, D. Marenduzzo, A. Maritan, and F. Seno, J. Phys. A **34** L751 (2001).

[37] I. D. Vladescu, M. J. McCauley, I. Rouzina, and M. C. Williams, Phys. Rev. Lett. **95**, 158102 (2005).

[38] L. Shokri, M. J. McCauley, I. Rouzina, and M. C. Williams, Biophys. J. **95**, doi:10.1529/biophysj.108.132688.

[39] J. R. Wenner, M. C. Williams, I. Rouzina, and V. A. Bloomfield, Biophys. J. **82**, 3160 (2002).

[40] J. Rudnick and T. Kuriabova, Phys. Rev. E **77**, 051903 (2008).

[41] H. J. Kreuzer, T. Einert, and R. R. Netz (unpublished).

[42] C. Storm and P. C. Nelson, Phys. Rev. E **67**, 051905 (2003).

[43] S. Cocco, J. Yan, J.-F. Leger, D. Chatenay, and J. F. Marko, Phys. Rev. E **70**, 011910 (2004).

[44] S. Whitelam, S. Pronk, and P. L. Geissler, *E-print* arXiv:0806.0505.

[45] S. Piana, Nucleid Acids Res. **33**, 7029 (2005).

[46] S. J. Rahi, M. P. Hertzberg, and M. Kardar, *E-print* arXiv:0806.2837.

[47] A. Hanke and R. Metzler, J. Phys. A **36**, L473 (2003).

[48] T. Ambjörnsson and R. Metzler, J. Phys. Cond. Mat. **17**, S1841 (2005).

[49] T. Ambjörnsson and R. Metzler, Phys. Rev. E **72**, 030901(R) (2005).

[50] T. Hwa, E. Marinari, K. Sneppen, and L.-H. Tang, Proc. Natl. Acad. Sci. USA **100**, 4411 (2003).

[51] H. C. Fogedby and R. Metzler, Phys. Rev. Lett. **98**, 070601 (2007)

[52] H. C. Fogedby and R. Metzler, Phys. Rev. E **76**, 061915 (2007).

[53] A. Bar, Y. Kafri, and D. Mukamel, Phys. Rev. Lett. **98**, 038103 (2007).

[54] R. D. Blake, R. D., Bizzaro, J. W., Blake, J. D., Day, G. R., Delcourt, S. G., Knowles, J., Marx, K. A., & SantaLucia, J., Jr. (1999) Bioinf. **15**, 370 (1999).

[55] M. Fixman and J. J. Freiere, Biopol. **16**, 2693 (1977).

[56] J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977).

[57] S. Caracciolo, M. S. Causo, and A. Pelissetto, Phys. Rev. E **57**, R1215 (1998).

[58] H.-P. Hsu, W. Nadler, and P. Grassberger, Macromol. **37**, 4658 (2004).

[59] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992).

[60] T. Ambjörnsson, M. A. Lomholt, S. K. Banik, and R. Metzler, Phys. Rev. E **75**, 021908

(2007).

[61] O. Krichevsky and G. Bonnet, Rep. Prog. Phys. **65**, 251 (2002).

[62] S. K. Banik, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **71**, 852 (2005).

[63] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell* (Garland, New York, 2002).

[64] C. G. Fry and P. J. Farnham, J. Biol. Chem. **274**, 29583 (1999).

[65] T. Ambjörnsson and R. Metzler, J. Phys. Cond. Mat. **17**, S4305 (2005).

[66] T. Novotny, J. N. Pedersen, M. S. Hansen, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **77**, 48001 (2007).

[67] M. E. Fisher, J. Stat. Phys. **34**, 667 (1984).

[68] J. N. Pedersen, T. Novotny, M. S. Hansen, T. Ambjörnsson, and R. Metzler (unpublished).

[69] R. L. Karpel, in A. Revzin, Editor, *The biology of non-specific DNA-protein interactions* (CRC Press, Boca Raton, FL, 1990).

[70] D. E. Jensen, R. C. Kelly and P. H. von Hippel, J. Biol. Chem. **251**, 7215 (1976).

[71] R. L. Karpel, IUBMB Life **53**, 161 (2002).

[72] K. Pant, R. L. Karpel, I. Rouzina, and M. C. Williams, J. Mol. Biol. **336**, 851 (2004).

[73] K. Pant, R. L. Karpel, I. Rouzina, and M. C. Williams, J. Mol. Biol. **349**, 317 (2005).

[74] T. Ambjörnsson and R. Metzler, Phys. Biol. **1**, 77 (2004).

[75] T. Ambjörnsson, M. Lomholt, and R. Metzler, J. Phys. Cond. Mat. **17**, S3945 (2005).

[76] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite (unpublished).

[77] Y. M. Wang, Robert H. Austin, and Edward C. Cox, Phys. Rev. Lett. **97**, 048302 (2006).

[78] J. Elf, G.-W. Li, and X. S. Xie, Science **316**, 1191 (2007).

[79] C. Venter and D. Cohen, New Perspectives Quarterly **21**, 73 (2004).

[80] R. J. Ellis and A. P. Minton, Nature **425**, 27 (2003).

[81] G. Rivas, F. Ferrone, and J. Herzfeld, EMBO Rep. **5**, 23 (2004).

[82] A. Kornberg, For the Love of Enzymes, (Harvard University Press, Cambridge, MA, 2000).

[83] I. Golding and E. C. Cox *Phys. Rev. Lett.* **96**, 098102 (2006).

[84] M. Weiss, M. Elsner, F. Kartberg, and T. Nilsson, Biophys. J. **87**, 3518 (2004).

[85] D. S. Banks and C. Fradin, Biophys J. **89**, 2960 (2005).

[86] I. M. Tolić-Nørrelykke, E. L. Munteanu, G. Thon, L. Oddershede, and K. Berg-Sorensen *Phys.*

Rev. Lett. **93**, 078102 (2004).

[87] M. Platani, I. Goldberg, A. I. Lamond, and J. R. Swedlow, Nat. Cell Biol. **4**, 502 (2002).

[88] Y. He, S. Burov, R. Metzler, and E. Barkai, Phys. Rev. Lett., at press.

[89] For an ideal chain embedded in d-dimensional space $d = 2d$ while for a self-avoiding walk the connectivity constant becomes reduced compared to that value. On average $\mu \approx 4.68$ in $d = 3$ [35].

[90] This relies on the assumption that $p_\ell(\mathbf{r})$ is *Gaussian* for ideal random loops and as described above Eq. (8) for self-avoiding loops. For very large βF denatured loops are stretched out and aligned along F so that the partition function is dominated by parameter values for which $p_\ell(\mathbf{r})$ deviates from this form. A suitable $p_\ell(\mathbf{r})$ should be used to obtain the phase diagram in this regime

[91] Due to intrachain coupling (e.g., Rouse), larger bubbles may involve an additional 'hook factor' $m^{-\mu}$ [48, 49].

[92] For diffusion time $\tau_D = 150\mu$s measured for an RNA construct of comparable length in [61].

[93] Other proteins are also involved in this quite complex process.

[94] In particular with the stability parameters from Krueger et al. [6] the stacking free energy of a `TA/AT` pair of base pairs is essentially vanishing.

[95] Due to the stabilisation of the coalesced bubble as discussed below this setup would allow for measurements of a first passage for the merging of the two initial bubbles and thus distinguish this setup from the open-close dynamics in the previous experiments.